Fax-On-Demand
Telephone: (202)401-0527
Item: 6026

A STATISTICAL METHOD FOR INCORPORATING NONDETECTED PESTICIDE RESIDUES INTO HUMAN HEALTH DIETARY EXPOSURE ASSESSMENTS
(Draft, 11/30/98)

Executive Summary

When residue data are submitted in support of establishing or reassessing a tolerance for a particular food use, in some cases a large portion of the data measurements of the levels of pesticide residue present on food shows no detection of residues. These "nondetects" (NDs) do not necessarily mean that the pesticide is not present at any level, but simply that any amount of pesticide present was below the level that could be detected using a particular analytical method.

The primary science policy issue concerning NDs is what value EPA should assign to them so that dietary exposure estimates will be as accurate and realistic as possible. Although the Agency generally assigns a default value of ½ the Limit of Detection (LOD) or Limit of Quantitation (LOQ) to NDs (see separate paper referred to below), EPA's policy would be to allow a person to use the statistical method described in this paper to more accurately determine the values or distribution of values for NDs. This method is fully described in EPA's *Guidance for Data Quality Assessment: Practical Methods for Data Analysis* issued in July 1996 (EPA/600/R-96/084), which has been peer reviewed by EPA program offices, regional offices and laboratories. This method would be used only in situations where the NDs comprise less than half the data set and the rest of the data are normally or lognormally distributed. Generally, the ND values obtained from this method would be less than ½ the LOD or LOQ but greater than zero.

In adopting this science policy, EPA's goal is to avoid underestimating exposure to potentially sensitive or highly exposed groups such as infants and children while attempting to approximate actual residue levels as closely as possible. Both biological information and empirical residue measurements support EPA's belief that this science policy is consistent with these goals. This paper also refers to two additional papers being made available for comment at the same time that address other aspects of the primary science policy issue mentioned above: (1) "Assigning Values to Nondetected/Nonquantified Pesticide Residues in Human Health Dietary Exposure Assessments", and (2) "Proposed Threshold of Regulation Policy Defining When a Food Use Does Not Require a Tolerance."[1]

Introduction

---

[1]These draft science issue papers are being made available for public comment for 60 days via an announcement in the Federal Register. EPA is seeking public comment on a series of draft documents concerning nine science policy issues in order to make the development of its FQPA-related science policies transparent and participatory for all interested parties.

Pesticide manufacturers (i.e., registrants) who petition EPA to establish a tolerance are required to submit data on the level of pesticide residues that remain in or on food. Often, instrumentation in the laboratory is not able to detect any residue below a specified level, which is called the "limit of detection" or LOD. However, even though the laboratory instrumentation cannot detect a residue, a residue may be present, at some level below the LOD, which may still present a potential concern to human health. Current EPA policy is generally to assume that non-detectable residues remain on treated commodities at ½ LOD. One issue that arises is whether the Agency's method for assigning finite values to NDs into its risk assessments may either overestimate or underestimate risk depending on the actual distribution of data below the LOD.

This science policy document describes statistical methods which may be used for determining the distribution of non-detectable residues below the LOD in cases where some of the residues of the data set are undetectable. When properly employed, such methods can provide a scientifically sound basis for more accurately estimating dietary exposure and risk than assuming that exposure occurs at ½ LOD or some other single, finite value. This document is intended to be used chiefly by persons conducting probabilistic human health dietary exposure assessments for purposes of registration or reregistration of pesticides. This guidance will help assure that dietary exposure assessments accurately portray exposures and risks to the U.S. population and subpopulations of special concern such as infants and children. Such assessments will play an increasingly important role in the evaluation of risks posed by pesticides and will improve the Agency's ability to make regulatory decisions that fully protect public health and sensitive subpopulations, including infants and children.

The Agency is separately issuing a draft science policy paper with respect to the use of "less than detection level" or "less than quantitation level" residues in dietary risk assessments. In this companion document, entitled "Assigning Values to Nondetected/Nonquantified Pesticide Residues in Human Health Dietary Exposure Assessments," a procedure is outlined which indicates under what conditions and which circumstances ½ LOD, ½ LOQ or other value should be assigned to samples reported as "nondetects." That policy paper states that other methods may, under certain circumstances, provide more accurate estimates of mean residue values for use in chronic exposure and risk assessments.

The purpose of the present document is to detail a proposed method for more accurately determining mean residue values from heavily "censored" data sets, i.e., data sets for which a substantial amount of data (e.g., 15-50%) are simply reported as less than a given detection or quantitation limit. The method discussed below (Cohen's method) is to be used only for data points which are part of a parent population which is normally distributed or that can be made normal via transformation. Practically, this means that the parent population should have either a normal or log-normal distribution. Prior to using this method, the existence of a normal (or transformed log-normal) parent population must be demonstrated. It is strongly recommended that the data be graphed on appropriate probability paper and that normality tests (e.g., Shapiro-Wilk) be performed to verify the assumed distribution. Various statistical procedures (with associated examples) which could be used to accomplish this task are available in the document

Guidance for Submission of Probabilistic Human Health Exposure Assessments to the Office of Pesticide Programs (draft dated 11/4/98) which is available for comment on the world wide web in draft format at http://www.epa.gov/fedrgstr/EPA-PEST/1998/November/Day-05/o-p29665.htm.

An additional criterion for use of this proposed methodology is that not more than 50% of the data set be censored (ideally, less than 20% should be censored). It is important to note that, when using USDA's Pesticide Data Program (PDP) or other monitoring data to calculate an average residue for use in a risk assessment, the percentage of the data set which represents "true zeroes" (i.e., not treated) should be eliminated from the data set before considering whether the procedure in this document is applicable. For example, if 80% of a crop is not treated, but 90% of the PDP values are reported as NDs, the bottom 80% of the data (from a rank-ordered set) should be removed from the data set (as they represent that portion of the crop which was not treated); the remaining NDs (i.e., 10% of the original sample) would be considered to represent treated commodities which have residues at levels lower than the LOD. Thus, in this case, 50% of the data would be censored (10% of the samples are ND and 10% of the samples are greater than the LOD).

Those measured values which lie above the LOD but below the LOQ should be considered as being "semi-quantitative." In contrast to the methodology described in the companion policy document entitled "Assigning Values to Nondetected/Nonquantified Pesticide Residues in Human Health Dietary Exposure Assessments" in which ½ LOQ is generally used as a default assumption for all values which lie between the LOD and LOQ, the actual measured "semi-quantitative" value should instead be used when working with methods for censored data (such as Cohen's method).

Cohen's Method

Cohen's method is fully described in EPA's *Guidance for Data Quality Assessment: Practical Methods for Data Analysis* issued in July 1996 (EPA/600/R-96/084) and in several other publications described in the Reference section of this document. The EPA publication is available on the Internet at http://Earth2.epa.gov/ncerqa/qa/docs/epaqag9.PDF. Briefly, the method uses the information provided by the uncensored portion of the data (i.e., that portion of the data with >LOD values) and the assumed normal (or transformed normal) distribution of the data to calculate a mean and standard deviation which incorporates the data which lie below the detection limit in the "censored" region of the data. Cohen's method requires that the distribution be normal (or can be made normal) and that there be only a single LOD or LOQ for all analyses of the same commodity. Since Cohen's method requires that the distribution be normal, the logarithms of the data must be used if the data are log-normally distributed with the resulting mean and standard deviation of the untransformed data calculated using the following formulae for the mean and standard deviation respectively:

$$M_a = \exp(M_L + 0.5\, s_L^2)$$

$$s_a^2 = M_a^2 \exp(s_L^2 - 1)$$

where M_a is the arithmetic mean of the original (untransformed) residue values, M_L is the mean of the logarithms of the residue values, s_L is the standard deviation of the logarithms of the residue values, and s_a is the standard deviation of the original (untransformed) residue values.

In general, the criterion that the data be normally (or lognormally) distributed is not expected to present an impediment to the widespread application of this technique. On a theoretical basis, concentrations of pesticides in food crops might be expected to be a Random-Product process and the Theory of Successive Random Dilutions (SRD) would predict that concentrations of pesticides would be lognormal (Ott, 1995). In addition, a fair amount of empirical evidence for a lognormal distribution of pesticides in foods exists from a recent study by the UK's Ministry of Agriculture, Fisheries, and Food (MAFF) in which thousands of individual serving sized samples were analyzed for a variety of pesticides and found to follow in many cases a lognormal distribution (MAFF, 1997).

Briefly, Cohen's technique for censored samples involves the following steps for log-normally distributed data (derived from Perkins, et al., 1990):

1. Determine N = total sample size
2. n = number of quantitated measurements
3. h = (N-n)/N
4. Transform the uncensored measurements to logarithms
5. Determine ln(LOD) = X_o
6. Determine $\underline{S}_L^2/(\underline{X}_L - X_o) = \gamma$ where $\underline{X}_L$ and $\underline{S}_L^2$ are the mean and variance of the log transformed detectable data, respectively.
7. Using appropriate tables (e.g., in US EPA, 1996 or Perkins) with h and γ, find λ
8. $M_L = \underline{X}_L - \lambda(\underline{X}_L - X_o)^2$
9. $s_L^2 = \underline{S}_L^2 + \lambda(\underline{X}_L - X_o)^2$
10. $M_a = \exp(M_L + 0.5\, s_L^2)$
    $s_a^2 = M_a^2 \exp(s_L^2 - 1)$

Estimation of Specific Values which Lie Below the Detection Limit

The Cohen method is appropriate for use in cases where it is sufficient to calculate a mean residue value and the basic distributional and other requirements are met. In general, the use of a mean value in a risk assessment is appropriate if a chronic analysis is being performed or if it is actually the mean value in an acute analysis which is of interest. In certain instances, it may not be

sufficient to simply obtain the mean (and standard deviation) of a data set by use of Cohen's method. For example, it may be desired to perform a Monte-Carlo analysis using data from a market basket survey in which single serving sized samples were analyzed and many NDs were obtained. Or it may be required to insert data from field trials with many ND values into a Monte-Carlo analysis. In these cases, it is the *entire set* of individual residue values (or their estimates) which are desired and not simply the mean or values which are greater than the LOD.

In these cases, it may be possible to use the information provided by the *uncensored* portion of the data to *impute* those values which lie below the detection limit in the "censored" region of the data.[2] As in the case with Cohen's method used to calculate a mean value, those data which are greater than the detection limit should be *demonstrated* to follow a specific distribution; it is this specific distribution which is used to extrapolate residue values into the less than detection limit (or censored) region of the data. The resulting imputed values can then be used directly in a Monte-Carlo assessment. This can be done in a number of ways:

1) the values that are above the detection limit can be plotted directly on the appropriate probability paper and the values at less than the detection level imputed directly via graphical analysis.

2) the values at less than detection levels can be imputed by graphical analysis as above, but could also be used instead to calculate the relevant defining parameters of the normal (or lognormal) distribution (i.e., the mean and standard deviation). This distribution, now fully defined, could be used to calculate the values less than the detection limit by using the inverse cumulative normal probability function available on many spreadsheets.

3) instead of using graphical techniques, the parameters of the normal distribution (mean and standard deviation) could be calculated by Cohen's method (described above). This now fully defined distribution could then be used to impute the values less than the detection limit by using the inverse cumulative normal probability function as in example 2.

---

[2] It is important that the non-treated NDs be removed from the distribution prior to performing statistical calculations (as discussed under the Cohen's Method section). Of course, one can instead simply substitute ½ LOD or ½ LOQ, as appropriate, for the NDs in a Monte-Carlo analysis.

Calculation of Average Residue Concentrations When No Distributional Assumptions Can Be Made

There are instances in which it is not possible to use Cohen's method or graphical analysis to better estimate residues at levels less than the Limit of Detection or Limit of Quantitation. For example, a market basket survey may reveal that only a small number of samples of a specific item (e.g., apples) contain detectable residues: in this case, one cannot establish with any degree of confidence a defined distributional pattern (e.g., lognormal) with those few data points which lie above the detection limit. Alternatively, data in USDA's Pesticide Data Program survey may have been developed in laboratories with several different (and potentially widely varying) limits of detection, thereby precluding the use of Cohen's method which needs a single (uniform) limit of detection (or quantitation). In these cases, the exposure assessor would substitute zeroes for residue values on crops which are not treated and ½ the LOD (or LOQ) for those crops which are treated but have ND residues. The overall methodology is described in the previously mentioned companion policy document. It is illustrated for further clarification in the box to the right with an example.

Suppose that 30% of apples are treated with a pesticide (and 70% are therefore not treated), but a PDP survey of 5 lb. composite samples shows that 80% (i.e., a total of 240 samples) of the 300 samples collected have ND (not detected or less than detection limit) residues. Three-quarters of those PDP ND values have a limit of detection of 0.05 ppm and one-quarter of the ND values have LOD's of 0.10 ppm. We wish to calculate the *average* residue in apples for use in a chronic dietary assessment.

Given this information, we would conclude that 70% of the 300 composite apple samples contain no (or zero) residues since they were not treated with pesticide. This means that 210 of the 300 composite samples are true zeroes (70%). From this, it follows that 210 of the 240 ND values (or 87.5% of the ND's) represent true zeroes with the remaining 30 ND values (or 12.5% of the 240 NDs) representing treated apples with residues at less than the detection limit. To calculate residues in these treated samples, we would assign one-half the 0.05 ppm LOD to three quarters of these ND's (representing an expected 22.5 of the 240 ND samples) and one-half the 0.10 ppm LOD to the remaining one-quarter of these ND's (representing an expected 7.5 of the 240 ND samples). The average residue for use in a chronic assessment would therefore be calculated as follows:

$$\frac{(210 \times 0 \text{ ppm}) + (22.5 \times 0.025 \text{ ppm}) + (7.5 \times 0.05 \text{ ppm}) + \sum (\text{all} > \text{LOD values})}{300}$$

We note that since composite (and not single-serving) samples were collected in this example, the risk assessor could NOT simply substitute the above values in a Monte-Carlo assessment in the appropriate proportions. It would only be appropriate to use the average residue (as illustrated above) in a chronic assessment.

Conclusion

A companion Federal Register announcement entitled "Assigning Values to Non-detected/Non-quantified Pesticide Residues in Human Health Dietary Exposure Assessments" describes a proposed science policy with respect to the use of "less than detection level" or "less than quantitation level" residues in dietary risk assessments. In those cases in which a data set contains a large percentage of non-detects (e.g., 15-50%) but the data set can be demonstrated to have a normal or log-normal distribution, it may be preferable to employ statistical methodologies to produce average estimators that better incorporate those values which lie below the detection

(or quantitation) limit.

In certain cases, one may instead desire to use the properties of a defined distribution to impute those values which lie below the detection limit and use these imputed values (in addition to those values greater than the detection limit) directly in a risk assessment. In these instances, a method is described which uses the known properties of a defined distribution to calculate those ND values which are not directly observable.

Finally, one may wish to calculate an average residue concentration when no distributional assumptions can be made as when, for example, very few >LOD values are present. In these instances, the methodology described in the companion, draft science policy paper mentioned earlier is available. It is illustrated in this paper for clarification.

References:

Berthouex, PM and LC Brown. 1994. *Statistics for Environmental Engineers*. Chptrs. 10 and 11. "Simple Methods for Analyzing Data that are Below the Limit of Detection" and " Estimating the Mean of Censored Samples", p.p. 81-95

Cohen, AC., Jr. 1959. Simplified estimators for the normal distribution when samples are singly censored or truncated. *Technometrics* **1**:217-237

Cohen, AC, Jr. 1961. Tables for maximum likelihood estimates: singly truncated and singly censored samples. *Technometrics* **3**:535-541

Helsel, DR. 1990. Less than obvious: statistical treatment of data below the detection limit. *Environ. Sci Technol*. **24**(12): 1766-1774.

Hornung, RW and LD Reed. 1990. Estimation of average concentration in the presence of nondetectable value. *Appl. Occup. Environ. Hyg.* **5** (1):46-51.

Ministry of Agriculture, Fisheries, and Food. Pesticides Safety Directorate. 1997. data supplied with *Unit to Unit Variation of Pesticides in Fruit and Vegetables*. March 14.

Ott, W. R. 1995. *Environmental Statistics and Data Analysis*. Chpts. 8 and 9 "Dilution of Pollutants" and "Lognormal Processes", p.p. 192-293.

Perkins, JL, GN Cutter, and MS Cleveland. 1990. Estimating the mean, variance, and confidence limits from censored (<limit of detection), log-normally distributed exposure data. *Am. Ind. Hyg. Assoc J*. **51**(8):416-19.

Travis, CC and ML Land. 1990. Estimating the mean of data sets with non-detectable values.

Environ. Sci. Technol. **24**(7):961-2.

U.S. EPA. 1996. Guidance for Data Quality Assessment: Practical Methods for Data Analysis EPA QA/G9. Office of Research and Development. EPA/600/R-96/084. (available in downloadable Adobe Acrobat pdf format on the world wide web at http://Earth2.epa.gov/ncerqa/qa/docs/epaqag9.PDF)

U.S. EPA. 1998. Draft document. Guidance for Submission of Probabilistic Human Health Exposure Assessments to the Office of Pesticide Programs. November 4. (available in downloadable Adobe Acrobat pdf format on the world wide web at http://www.epa.gov/fedrgstr/EPA-PEST/1998/November/Day-05/o-p29665.htm)